INSTITUTE FOR DEFENSE ANALYSES

Regularization for Continuously Observed Ordinal Response Variables with Piecewise-Constant Functional Predictors

Matthew Avery
Laura Freeman
Timothy Robinson
Mark Orndorff

January 2016

IDA Document NS D-5699
Log: H 16-000002

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

This paper investigates regularization for continuously observed covariates that resemble step functions. The motivating examples come from operational test data from a recent United States Department of Defense (DoD) test of the Shadow Unmanned Air Vehicle system. The response variable, quality of video provided by the Shadow to friendly ground units, was measured on an ordinal scale continuously over time. Functional covariates, altitude and distance, can be well approximated by step functions. Two approaches for regularizing these covariates are considered, including a thinning approach commonly used within the DoD to address autocorrelated time series data, and a novel "smoothing" approach, which first approximates the covariates as step functions and then treats each "step" as a uniquely observed data point. Data sets resulting from both approaches are fit using a mixed model cumulative logistic regression, and we compare their results. While the thinning approach identifies altitude as having a significant impact on video quality, the smoothing approach finds no evidence of an effect. This difference is attributable to the larger effective sample size produced by thinning. System characteristics make it unlikely that video quality would degrade at higher altitudes, suggesting that the thinning approach has produced a Type 1 error. By accounting for the functional characteristics of the covariates, the novel smoothing approach has produced a more accurate characterization of the Shadow's ability to provide full motion video to supported units.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-5699

Regularization for Continuously Observed Ordinal Response Variables with Piecewise-Constant Functional Predictors

Matthew Avery
Laura Freeman
Timothy Robinson
Mark Orndorff

Regularization for Continuously Observed Ordinal Response Variables with Piecewise-Constant Functional Predictors

1. Introduction

Many engineering applications involve the collection of functional data in which the unit of measurement is a curve measured over a continuous time domain. In this manuscript, we consider a case study involving a Follow-On Operational Test and Evaluation (FOT&E) conducted by the United States Army on the RQ-7BV2 Shadow, a tactical unmanned air vehicle system tasked with, among other tasks, providing full motion video to supported ground units. The quality of full motion video was the response variable of interest, and it was recorded on an ordinal scale continuously over the duration of each mission.[1] In addition, important covariates, such as the air vehicle's altitude and the distance from the air vehicle to the supported unit, were also observed continuously over time. With respect to notation, we let $Y_i(t_{ij})$, $X_{1i}(t_{ij})$, and $X_{2i}(t_{ij})$ be functions describing the video quality, altitude, and distance from the air vehicle to the supported unit, respectively, for mission i, measured at time t_{ij} ($1 \leq i \leq n = 54; 1 \leq j \leq m_i$), where $t_{ij} \in T_i$, denotes the time domain for the ith mission. The value n denotes the total number of missions from the test and the subscript m_i denotes the number of observations (i.e. time points) for mission i. While functional data is common in practice and the field of functional data analysis has progressed in recent years [see Ramsay and Silverman (2005) for an overview of these analytical methods], there has been little if any research regarding the analysis of *function-on-function* modeling involving an ordinal response and continuous covariates.

Functional regression problems are generally categorized into three types:

1. Functional predictor regression in which the response variable is regressed on functional covariates
2. Functional response regression in which a functional response variable is regressed on scalar covariates
3. Function-on-function regression in which a functional response is regressed on functional covariates.

Regardless of the type of functional regression problem, there is a need to combine information both across (replication) and within (regularization) functions. In this context, we refer to "replication" from sampling perspective rather than the experimental design perspective. Like most DoD testing, our example contains no pure replication with factors like altitude and distance being fixed. Data are observed in the following triplicate for each mission: $[Y_i(t_{ij}), X_{1i}(t_{ij}), X_{2i}(t_{ij})]$. Information in these triplicate observations cross the $i = 1, 2, \ldots, n = 54$ missions is combined to construct a predictive model that describes the relationship between

[1] Each mission consists of the period of time between the air vehicle's launch and recovery during which the air vehicle was tasked to provide full motion video to a ground unit participating in the test, such as the First Battalion, 6th Infantry Regiment.

the response and covariates. In our example, the predictive model will enable the Department of Defense (DoD) to profile video quality given the Shadow's altitude and distance. Regularization involves imposing constrains on the data collected over time in order to borrow information across time periods (i.e., observations "within" a function) that are expected to be similar to each other. If a variable is expected to have similar values across a range of time, then the estimation of the function at time *t* is improved by borrowing the information on this variable from nearby observations in time. For many functional data problems, regularization involves some type of data smoothing. Regularization can also be used to prevent overfitting by accounting for potential redundancy of time points that are similar to one another.

In this paper, we will discuss the analysis of a function-on-function problem and compare the use of functional smoothing to an alternative approach that focuses on thinning. The thinning approach is commonly used in many communities including the DoD when data are highly auto-correlated. In the comparison, we will illustrate problems with thinning when the functional structure of the data is more complex than a simple autocorrelation.

This paper is motivated by a desire to develop a predictive capability linking the covariates distance and altitude to the video quality of the Shadow UAV. To exploit information across missions in a predictive analysis, we use an ordinal mixed regression with both approaches. The remainder of the paper is organized as follows. Section 2 summarizes the RQ-7BV2 Shadow, the test that generated the data, and characteristics of the data itself. Section 3 discusses the two regularization approaches considered to prevent overfitting and leverage information across time within each mission. Section 4 provides background on cumulative logistic regression and generalized linear mixed models, which are applied to the both of the regularized datasets. Section 5 compares results from these two methods discussed, and Section 6 contains concluding remarks.

2. Operational Testing of the Shadow UAV

Prior to fielding new capabilities, US Army systems undergo testing meant to mimic the real environment in which the system will be employed. This process, called operational testing, is a key part of the United States Department of Defense (DoD) acquisition process. These tests use systems that are representative of the versions that will be deployed (e.g., production line systems as opposed to prototypes) and are conducted in environments similar to those in which the system will be used. System operators have military rank, training, and experience consistent with the system's concept of employment. The goal is to characterize the capabilities of the system that the warfighter can expect should the system be used in combat.

The Network Integrated Evaluation (NIE) in a twice-annual event conducted by the US Army to evaluate and help improve integration of the Army's tactical communication effort. These events include hundreds of US Army soldiers, often including units preparing to deploy overseas, including a brigade-sized unit of friendly "blue" forces simulating a multi-week fight against an opposing "red" force. The iteration in May of 2014 presented a good opportunity to exercise the

full capabilities of the Shadow and observe how a US Army brigade was able to employ it in a simulated combat environment. Of particular interest at this test were upgrades to Shadow's data link (through which the air vehicle receives commands from the ground and sends back imagery) and interface with ground units. Over the course of the test, 54 missions were recorded, including over 244 flight hours.

The Shadow[2] is a tactical unmanned air vehicle system with the primary mission[3] of reconnaissance, surveillance, target acquisition, and battle damage assessment. It is operated by two Soldiers in a ground control station, one who flies the air vehicle and the other who operates the payloads, including both optical and infrared cameras capable of capturing full-motion video. The aircraft has a wingspan of over 20 feet, weighs roughly 460 pounds, is launched pneumatically, and normally lands on a paved runway. It has already seen over 970,000 flight hours of use by the US Army in both Iraq and Afghanistan. Typical missions include investigating particular areas in search of enemy units, reconnaissance of convoy routes to detect enemy activity, and surveillance of high-value targets or buildings. Forward-deployed units supported by the Shadow may be equipped with remote video terminals capable of receiving full-motion video from the Shadow payload in real-time. Analysis of the Shadow's capability to provide continuous full-motion video to supported units as a function of the Shadow's altitude and distance provides the motivation for this paper.

A typical mission for Shadow might proceed as follows:

1. After launching, the Shadow receives tasking to investigate a particular area in support of one or more ground units capable of receiving video transmissions from the Shadow.
2. An airspace coordination cell will assign the Shadow air space in which to operate, and the Shadow will proceed towards the area, staying within the altitude and location restrictions of its assigned airspace.
3. Once it arrives on station, the Shadow will maintain altitude with a circular orbit, typically off-set from the target or area of interest.
4. As the needs of the supported unit change, the Shadow's tasking is altered dynamically, often requiring the Shadow to re-position to accommodate the new tasking.
5. The Shadow is re-tasked as needed throughout the duration of its mission (potentially up to 9 hours) until its mission is complete and it returns to base.

Over the course of a mission, Shadow will change its location in response to new assignments and may change altitude due to airspace restrictions that change as other air assets (helicopters, etc.) enter and leave the area. Additionally, Shadow may climb to provide better line of sight to

[2] Though this paper focuses on the capabilities of the air vehicle, the RQ-7BV2 Shadow system also includes ground equipment, such as the pneumatic launcher, recovery system, and the equipment by which the air vehicle is controlled from the ground. Since these components are not of interest in this paper, we will use the terms "Shadow" and "air vehicle" interchangeably.

[3] Shadow has many other capabilities and missions that were evaluated during this test event, but our data and the unique challenges that came with it concern the ISR capabilities.

supported units or descend to avoid cloud cover, which can prevent the camera from observing activity on the ground.

Summary of Test Data

Data resulting from this test have unique characteristics. For each mission, the quality of the video provided by Shadow is recorded as a three-level[4] ordinal variable. This data was collected by Army officers co-located with the supported units. These officers observed the video provided by the Shadow in real-time, recording the initial quality of the data on the ordinal scale and noting the times at which video quality degraded or improved. The resulting timeline captures the video quality at any point during the mission, rated ordinally. Shadow's altitude and distance were recorded over time at a rate of $\frac{1}{60}$ Hz (once per minute). Figure 1 plots these three variables for a representative mission.

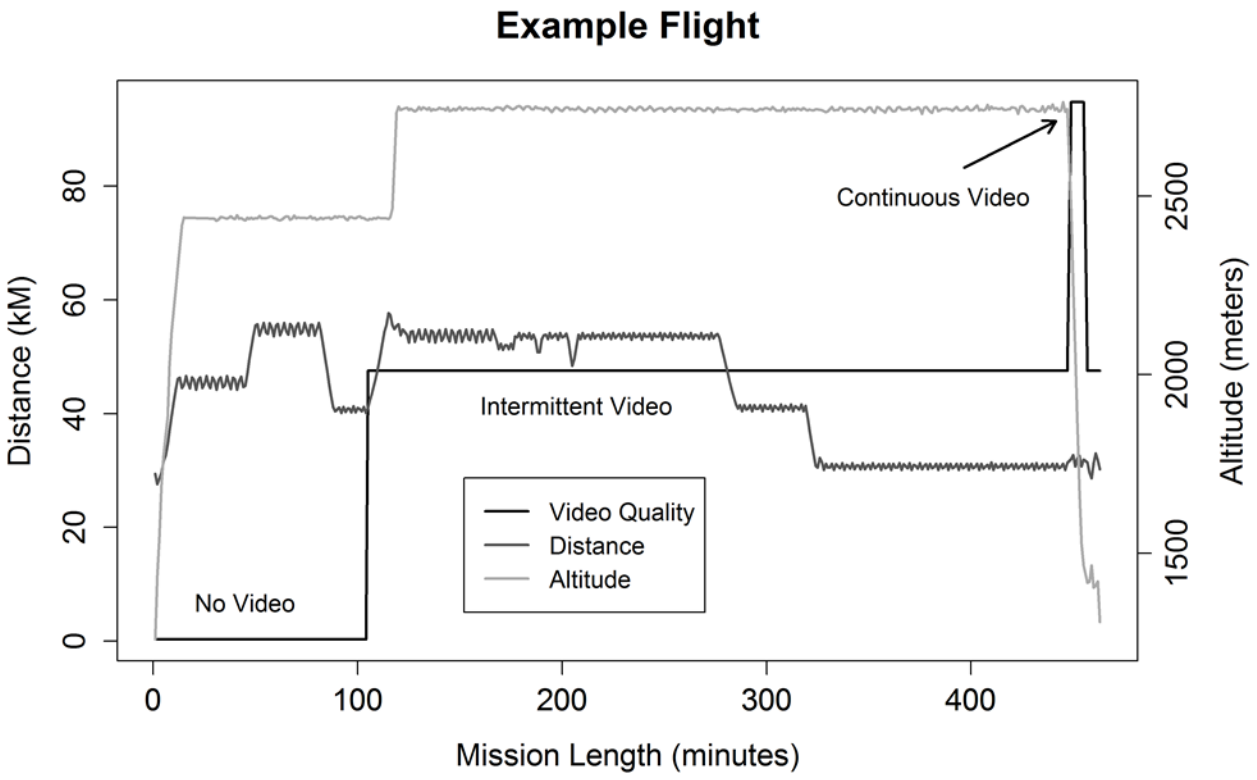

Figure 1: Example flight from RQ-7BV2 Shadow FOT&E. Video quality is measured on a three-level ordinal scale, increasing from "No Video" to "Intermittent Video" to "Continuous Video".

Since video quality is observed ordinally, its profile over time presents itself as piece-wise constant for periods of time. Both covariates, distance and altitude, also have piece-wise constant

[4] No Video, Intermittent Video, and Continuous Video

profiles over time with small perturbations that exhibit high autocorrelation. The small perturbations in distance are due to the circular orbit of the air vehicle. Relative to both the magnitude of the distance and the larger changes in distance (ex. Note the small perturbations in distance from 320 minutes to 450 minutes), this cyclical variation is negligible. Based on the behavior of the air vehicle described above, it is reasonable to think of these covariates as step functions: Changes in altitude are infrequent during most missions, and distance to target tends to change substantially only when the air vehicle is repositioning for a new task. The median first order autoregressive correlation across all missions is r=0.85, but this does not tell the whole story. Autocorrelation is highly variable within missions. Over regions where Shadow maintains altitude and is not re-tasked, both covariates will have autocorrelations near 1, since Shadow did not change position beyond its orbit. As Shadow repositioned, both altitude and distance changed substantially in short periods of time, yielding much lower autocorrelations during those periods. In the following sections, we will discuss two approaches for addressing the autocorrelation and piece-wise constant structure of these variables as a possible methods for regularizing the data.

3. Regularization Approaches for Individual Missions

As noted earlier, video quality, altitude and distance exhibit piece-wise profiles across time and hence much of the data observed in real-time is redundant in nature. In most functional data applications, information across time that is observed to be similar (and hence exhibits high auto-correlation) is distilled into a summary measure. This process is known as regularization. In this section, we discuss two approaches to regularizing Shadow data.

Thinning Approach

One common approach to addressing autocorrelation among observations within a given mission is to thin the data. Thinning is often used in the DoD testing community to account for autocorrelated data observed over time and has been applied to a disparate systems from radars to UAVs. While thinning is not typically considered a regularization approach by the functional data community because it ignores complex functional structures, it is relevant to our analysis because of its prevalence in use and the ease implementation.

Generally, thinning data by a rate of λ means sampling the original data set such that a data set of size n (rounded to the nearest integer) is generated, where n is the total number of data points in the autocorrelated data set and $\lambda \in (0,1)$. By letting $\lambda = 1 - \sigma_a$ where σ_a is an estimate of the autocorrelation in our data set, we can theoretically sample data in such a way that they are independent by taking every hth observation, where $h = \left\lceil \frac{1}{\lambda} \right\rceil$ [5]. However, to achieve independence the autocorrelation must be constant across the full time range.

[5] The ceiling function.

Looking to the right side of Figure 1 (near t=447), we can see that the response variable can change quickly at times. Should the thinning be performed systematically as described above, such periods may be missed completely. To avoid this scenario, we drew $(1-\lambda)\sum_{i=1}^{n} m_i$ observations from the $N = \sum_{i=1}^{n} m_i$ total number of data points without replacement. The autocorrelation from the Shadow data set across all missions was $\lambda = 0.85$, so only 15% of the data was retained post-thinning. After thinning the data, we achieve an effective sample size of 1,983 data points or 15% of the 13,496 observations taken at one minute intervals. As will be discussed in more detail in Section 5, a cumulative logistic mixed model regression was fit to the thinned data, and factor effects for altitude and distance were estimated. To account for the stochastic variability of this approach, this process was repeated 500 times. Parameter estimates and associated standard errors were recorded, and the median values across the 500 runs was used to compute a p-value for the significance of each factor. The results from this approach are shown in Section V. Table 1 shows a subset of the 23 data points sampled in the 500th iteration via thinning from the example mission.

Functional Smoothing Approach

The thinning approach ignores several features of the data, including the non-constant autocorrelation within missions. By using thinning to regularize the data, have reduced the amount of information available when fitting our regression model in proportion to the autocorrelation observed across the whole data set, but this brute force approach ignores the unique structure of each mission. For example, a mission with 10 different distance (location) changes and no altitude changes provides different information than a mission with 2 distance (location) changes and two altitude changes. Thinning assumes that each observation is correlated with observations adjacent in time, and that this correlation is constant over time. Contrary to this assumption, the data shown in Figure 1 could more accurately be described as piece-wise constant curves. Both distance and altitude remain constant for long periods of time, then exhibit rapid changes, then return to being constant or near-constant. This behavior can be accurately modeled using step functions. The distance factor resembles a step function contaminated by small cyclical perturbations, with the perturbations resulting from the circular loiter pattern of the air vehicle when holding position near a target or area of interest. Since these perturbations are small in magnitude relative to changes in distance that occur when the air vehicle is assigned a new task, the loss of this feature when approximating the factor as a step function should have minimal impact on our results. The approach we present below accounts for the structure of the observed curves while still allowing us to fit a regression model for prediction and variance quantification.

To model altitude and distance as step functions, we approach the data on a mission-to-mission basis. The critical challenge is to accurately identify time points corresponding to changes in the covariates. The change points for video quality are already recorded in the data. Using the set of change points from both covariates and the response variable, we can estimate each factor's value over the regions between each change point. Over these regions, all three curves should be

constant or near-constant, making estimation of the functions for each region trivial. This process is summarized in the following list:

1. Identify change points for altitude and distance for each mission.
2. For each mission, generate a set of change points by combining the change points from altitude, distance, and video quality for that flight.
3. Estimate the value of each factor for each interval between two change points. We used the median over the range to protect against sensitivity to the exact placement of the change point.

The result of this process is a step function estimate of each factor for each flight.

Change Point Identification

In many applications, identifying change points for a given curve requires both identifying the number of change points and identifying the correct locations in time where the changes occur. These locations are often referred to as knots[6] in the literature on spline fitting. As mentioned in Jupp (1975), finding the proper locations is computationally difficult for larger datasets with a large amount of locations. In many cases, selecting the number and locations of knots can be done by hand. This is undesirable in our case and others for several reasons. First, since the number of functions to be approximated is large, selecting knot locations by hand will be time consuming. Second, human error is introduced when deciding proper knot placement. Though still computationally costly, modern computational capabilities and algorithms make automated knot-fitting a viable option for our data set.

The literature contains several methods for identifying knot locations. Processes we considered can generally be described as having two steps. First, the algorithm will find the optimal location for a fixed number of knots. Next, cross validation method is used to compare fits using various numbers of knots to select the optimal number of knots. One method, proposed in Picard et al. (2005) assumes that the data are generated by a Gaussian process with the mean specified by a step function and independent variances within a cluster. Dynamic programming is used to reduce the computational complexity of the algorithm to $O(n^2)$. Another approach is to use a Haar wavelet basis with even/odd cross validation, as proposed by Nason (1996). The Haar basis is piecewise constant, making it well suited to approximating a step function. The advantage to this method is that the wavelet decomposition is computationally fast, allowing it to work well when other methods may prove to be computationally infeasible. Though we did not use this approach, Haar wavelets should be investigated in future work as a potential alternative to the approach described below.

[6] In this manuscript, we use the phrase "knots" exclusively in the context of spline fitting. We include this note to avoid confusion among readers familiar with aviation or nautical terminology, where the same word is used as a measure of speed.

The approach used in this application was to approximate the step function via a spline of degree 0. Using the methods of Spiriti et al. (2013), a genetic algorithm was used to find the optimal knot locations. The optimal locations are those which minimize the residual sums of squares between the approximation and the data collected. Denoting the covariate to be approximated by X_{ki}, the residual sums of squares criterion for r knots for the kth factor on the ith mission is given by

$$\mathrm{RSS}(\boldsymbol{b_i}, \boldsymbol{\xi_i}) = \sum_{j=1}^{m_i} \left(X_{ki}(t_{ij}) - \sum_{l=1}^{r_i+1} b_{il} B_{il}(t_{ij}; \boldsymbol{\xi_i}) \right)^2,$$

where $\boldsymbol{\xi_i}$ denotes the set of knot locations, B_{il} a spline function of degree zero, and b_{il} the coefficient for the lth spline function for the ith mission. The criterion was then calculated for each number of possible knots $\in 1, \ldots, r_{max}$.

Generalized cross validation (GCV) reduced the computational time of this process. The number of knots must be identified for each covariate for each mission, meaning the computationally taxing process described above must be repeated many times. To reduce overall computation time, generalized cross validation (GCV) was used. Since the change points from both covariates and the response variable will be combined for each mission, it was desirable to identify a parsimonious set of knots that is adequate for modeling each covariate. Therefore, we added a penalty term based on variance of the covariate. Denote the number of knot locations selected by GCV for a single covariate to be by $\hat{r}_{GCV} = \underset{\mathrm{r} \in 1,2,\ldots \mathrm{r}_{max}}{\arg\min}\ \mathrm{GCV}(r)$. Instead, we choose the largest $\hat{r}$ such that $\mathrm{GCV}(\hat{r}) \leq \mathrm{GCV}(\hat{r}_{GCV}) + \mathrm{SD}(\mathrm{GCV}(\boldsymbol{r}))$ and $\hat{r} \leq \hat{r}_{GCV}$. As the number of knots increases, $\mathrm{GCV}(r)$ decreases, with the rate of decrease eventually approaching 0. Penalizing based on the standard deviation of $\mathrm{GCV}(r)$ pushes $\hat{r}$ towards the "elbow" of the curve defined by $\mathrm{GCV}(r)$ and away from the tail. While this penalty was shown to improve the function estimates for this data set, it may not be necessary for other applications.

Step function approximation

Once the knots are found, we combine the change points from all factors (the observed change points from video quality and the estimated knots from altitude and distance) for each mission. For mission i, let $\hat{\xi}_1^{1i}, \hat{\xi}_2^{1i}, \ldots, \hat{\xi}_{k_i^1}^{1i} \in \hat{\boldsymbol{\xi}}^{1i}$ be the set of knots for altitude. Similarly, let $\hat{\boldsymbol{\xi}}^{2i}$ and $\boldsymbol{\xi}^{3i}$ be the set of estimated knots for distance and observed change points for video quality for mission i respectively. Then $\hat{\boldsymbol{\xi}}^i = \hat{\boldsymbol{\xi}}^{1,i} \cup \hat{\boldsymbol{\xi}}^{2,i} \cup \boldsymbol{\xi}^{3,i}$ is the estimated set of all change points for mission i. To eliminate redundancy, change points that are close together in time (less than five minutes apart) are combined. Figure 2 shows the change points for each variable for our example mission. The knots from the estimated altitude and distance curves as well as the change points recorded for video quality have been shown in dotted lines on each of the charts. Vertical lines represent estimated change points, with the symbol used in the line indicating the factor

generating the change point. If we denote our example flight as mission 7, then the line of "x"s represents the single knot in $\hat{\boldsymbol{\xi}}^{11}$, the lines of "+"s represent the knots from $\hat{\boldsymbol{\xi}}^{21}$, and the lines of "◊" represent the change points from $\boldsymbol{\xi}^{31}$. Taken together, these change points along with the beginning and end of the mission define a set of $r_1 + 1 = 8$ intervals on the domain of this flight, T_1. Let $\xi_1^1 < \cdots < \xi_{r_1}^1 \in \boldsymbol{\xi}^1$ be the change points for our example mission, and $\tau_l = (\xi_l, \xi_{l+1}]$ be the intervals between each change point. Then $\cup_{l=1}^{r_1+1} \tau_l = T_1$. Finally, we must estimate values of the step function for each interval. For altitude, let $\hat{X}_{11}(t) = c_l^1 \; \forall \, t \in \tau_l$. Since the curves are typically near-constant over these intervals, the simple median is adequate. Thus, $c_l^1 = median\{X_1(t) | t \in \tau_l^1\}$.

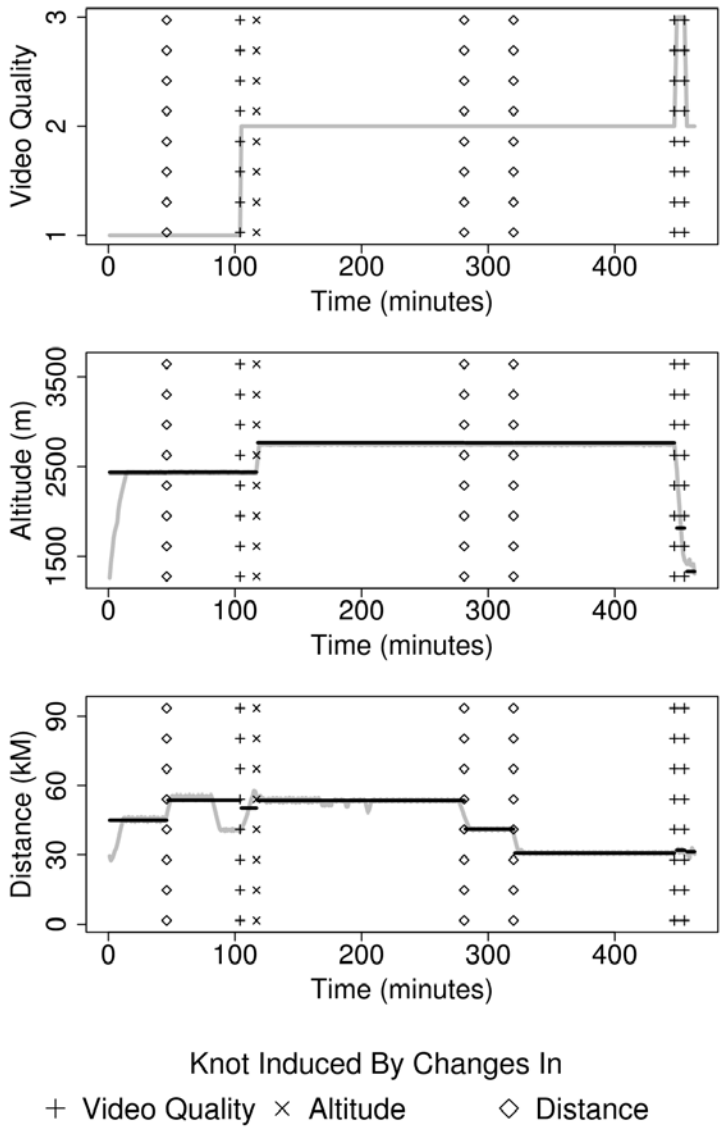

Figure 2: Observed values and fitted 0-order splines (for altitude and distance) with change points included from each function.

This representation of the data is simultaneously simpler and consistent with the complex functional nature of the data. Table 2 shows our example mission represented in tabular form. Applying the same procedure to all 54 missions in our data set yields a data matrix with a total of 347 rows, with each row corresponding to a period of time during a flight when the functional covariates and response were constant. Note that the effective sample size resulting from this approach (347 effective observations) is substantially smaller than the sample size resulting from the thinning regularization (1,983 effective observations).

4. Predictive Modeling Approach

After performing the regularization of the data within each mission, we next need to construct a predictive model that links the information across missions to determine the effect of the distance and altitude covariates to the video quality response variable.

Cumulative Logistic Regression

Proportional odds logistic regression is a standard approach for modeling independent ordinal responses, and the time dependent nature of functional data can be addressed via a mixed ordinal logistic model. Before discussing the mixed ordinal logistic model, we will review the ordinal logistic model for independent responses. For a response variable y taking on ordinal values 1 to C and a 1xk vector of explanatory variables, $\boldsymbol{x}_{ij}$, the proportional odds model is given by

$$\log\left(\frac{F_{ic}}{1-F_{ic}}\right) = \alpha_c + \boldsymbol{x}_i\boldsymbol{\beta}, \qquad c = 1, \dots, C-1 \tag{1}$$

where $\boldsymbol{x}_{ij}\boldsymbol{\beta} = \beta_1 X_1 + \cdots + \beta_k X_k$ and F_{ic} denotes the probability that the response for unit i falls into the c^{th} category or lower. More explicitly,

$$F_{ijc} = P\left(Y_{ij} \le c\right) = \frac{\exp(\alpha_c + \boldsymbol{x}_{ij}\boldsymbol{\beta})}{1+\exp(\alpha_c + \boldsymbol{x}_{ij}\boldsymbol{\beta})}, \quad c = 1, \dots, C-1.$$

This model is also referred to as the cumulative logit model, since the explanatory variables predict the cumulative probabilities of y. Other common ordinal logistic regression models include the adjacent-category model and the continuation-ratio model [see Agresti (2007), and Agresti (2012)]. We utilize the cumulative logit model in our analysis, but the methodology presented here could also be applied with other link functions. Therefore, for our application, let $\boldsymbol{x}_{il}$ be the lth regularized data point from the ith mission. Then

$$F_{ilc} = P\left(Y_{ij} \le c\right) = \frac{\exp(\alpha_c + \boldsymbol{x}_{il}\boldsymbol{\beta})}{1 + \exp(\alpha_c + \boldsymbol{x}_{il}\boldsymbol{\beta})}, \qquad c = 1,2.$$

Generalized linear mixed models

In many applications, data is observed longitudinally in clusters. With respect to this application, observations occur across time for each mission and the mission represents a cluster of

observations. Missions are assumed independent of one another but observations across time within a given mission are inherently correlated, and the correlation structure must be properly accounted for in the analysis. For multilevel data such as this, random cluster (ex. here cluster = mission) effects can be added into the regression model to account for this correlation. The resulting model is a *mixed* model containing *fixed* or *population averaged* effects which address systematic variation across the missions and *random* or *subject-specific* effects which address within mission variation. Mixed models for continuous normal outcomes were first presented by Laird and Ware (1982), and these models appear extensively in the literature. In non-normal data situations, these mixed models are commonly referred to as *generalized linear mixed models* (GLMM) and Myers, Montgomery, Vining and Robinson (2010) discuss applications of these models to engineering and industrial data.

Assuming there are $i = 1, \dots, n$ independent missions with $l = 1, \dots, r_i$ regularized observations per mission, the GLMM relates the conditional mean for the i^{th} mission to the fixed and random effects as follows,

$$E[Y_{il}|\boldsymbol{\delta}_i, \boldsymbol{x}_{il}] = g^{-1}(\boldsymbol{x}_{il}\boldsymbol{\beta} + \boldsymbol{z}_{il} + \boldsymbol{\delta}_i),$$

where Y_{ij} is the response in mission i at the lth regularized data point, g is a differentiable monotonic link function, $\boldsymbol{x}_{il}$ is the associated (1x2) vector of fixed effect model terms (altitude and distance), $\boldsymbol{\beta}$ is the corresponding (2 x 1) vector of fixed effect regression coefficients, $\boldsymbol{\delta}_i$ is the (q x 1) vector of random factor levels associated with the i^{th} mission and $\boldsymbol{z}_{ij}$ is the (1 x q) vector of random effects. For the thinning approach, there are $(1-\lambda)\sum_{i=1}^{n} m_i = N_{thin} = 1{,}983$ effective data points per sampled itteration, and for the smoothing approach, there are $\sum_{i=1}^{n}(r_i + 1) = N_{smooth} = 347$ effective data points. The conditional response, $E[Y_{il}|\boldsymbol{\delta}_i, \boldsymbol{x}_{il}]$, is assumed to have an exponential family member distribution and each of the q random effects are assumed normally distributed with mean zero, and the variance-covariance matrix of the vector of random effects in the i^{th} mission is denoted $\boldsymbol{D}_i$. The $\boldsymbol{D}_i$ are typically taken to be the same for each cluster.

Extending the cumulative logit model to the GLMM, the conditional cumulative probability of Y_{ij} resulting in an outcome in the c^{th} category can be denoted as $F_{ijc} = P(Y_{ij} \leq c|\boldsymbol{\delta}_i, \boldsymbol{x}_{ij})$. The GLMM representing the probability the response for unit i at time j falls into the c^{th} category or lower is given by

$$F_{ijc} = P(Y_{ij} \leq c|\boldsymbol{\delta}_i, \boldsymbol{x}_{ij}) = \frac{\exp(\alpha_c + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\delta}_i)}{1 + \exp(\alpha_c + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\delta}_i)}$$

where the notation is as defined previously.

5. Comparison of Results

Data from both regularization approaches (i.e. thinning and smoothing) were analyzed using a cumulative logistic mixed model regression, as discussed in Sections IV. The random effect for mission was statistically significant (p-value <0.01) for both approaches, and no second-order interaction effects were significant. For the thinning approach, five hundred Monte Carlo data sets were. Table 3 shows the median parameter estimates and median standard errors across the 500 runs with associated p-values. Table 4 shows results and p-values from the cumulative logistic mixed model based on the data set generated by smoothing. Due to concerns over operational security, the altitude and distance values were transformed to obscure the exact impacts these factors have on system performance.

Both sets of results report that video quality tends to degrade the further away from the supported unit the air vehicle is. The negative parameter estimates for altitude resulting from both regularization approaches are counterintuitive. Since the range of altitudes observed during this test (1,200 to 3,600 meters) are small relative to the horizontal distances between the air vehicle and supported unit, differences in altitude are unlikely to make a substantial impact on the strength of the signal from the air craft to the supported unit. Therefore, the principal impact one might expect from altitude would be on the line-of-sight between the air vehicle and the antenna capturing the video for the supported unit. If this line-of-sight is obstructed by brush, small trees or other such microterrain, video quality may degrade. Aircraft flying at higher altitudes are less likely to face line-of-sight obstructions, so if altitude were to have any effect on video quality, we would expect improved performance at higher altitudes. The date of test factor is included to model changes in performance over the duration of the test. The positive parameter estimate for Date of Test indicates that Shadow performance improved over the duration of the exercise, perhaps as crews became more familiar with operating the air vehicle and device used to capture video broadcast by the air vehicle.

One major difference in the results from the two methods is the statistical significance of altitude. Though the parameter estimates are similar, the model fit to the data generated using the step function approximation of the factors shows a larger standard deviation and a much larger p-value than the thinning approach generated. Within the DoD, p-values smaller than 0.2 are typically regarded as statistically significant. By this standard, using the thinning approach would lead us to conclude that altitude had a significant negative effect on video quality. We believe this conclusion is incorrect based on the results of the step function approach which incorporates the piece-wise constant structure of the altitude factor. This conclusion is also supported by our *a priori* knowledge of the system, which suggests that video provided by Shadow should not degrade in response to changes in altitude.

The difference in reported p-values is attributable to the difference in standard errors reported by these methods. Note that the standard error reported based on the smoothing regularization is much larger than the standard error based on thinning. This is due in part to the much larger effective sample size (1,983 vice 347) that results from thinning vice smoothing. As data are

collected on an increasingly granular scale, it is vital that appropriate methodologies are used to regularize the data and accurately represent the true amount of information.

Understanding population-level system performance as well as mission-to-mission variability is critical, and our approach provides estimates for both. Figure 3 shows the population estimates of the probabilities of experiencing varying levels of video quality as a function of the distance from the air vehicle and supported unit midway through the test (date of test = 11). The lower black line show the estimated probability of experience video quality of at least Intermittent Video or higher, and the higher black line shows the probability of experiencing full motion video. The cumulative probabilities were computed by averaging over the random effects.

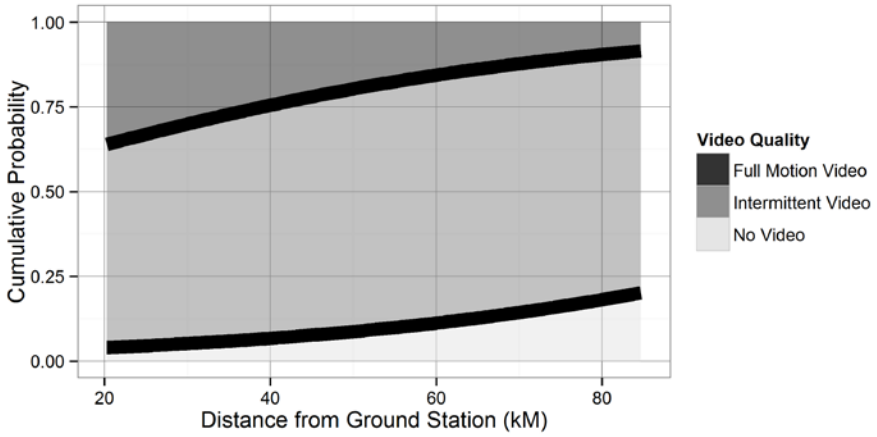

Figure 3: Cumulative probability of video quality by distance from ground station

Looking only at population-level estimates obscures the considerable mission to mission variation observed during testing. Figure 4 shows the estimated video quality thresholds for each mission in gray lines with the thinker black line representing the population estimate shown in Figure 3. The left panel of Figure 4 shows the probability of no video as a function of distance, and the right panel shows the probability of no video or intermittent video. There is considerable variability from mission to mission. For example, the probability of intermittent video or no video (right panel) on the best mission was lower than the population mean probability of no video(left panel). Mission-to-mission variability can be attributed to any of a large number of sources, potentially including meteorological conditions during the mission, the air vehicle and support equipment used for the mission, the proficiency of the air vehicle crew, and the proficiency of the soldiers in the supported unit with the video capture device. Since all of these factors may also vary when the Shadow system is deployed, quantifying both the population average and mission-to-mission variability provides the US Army with valuable information on the utility of the system.

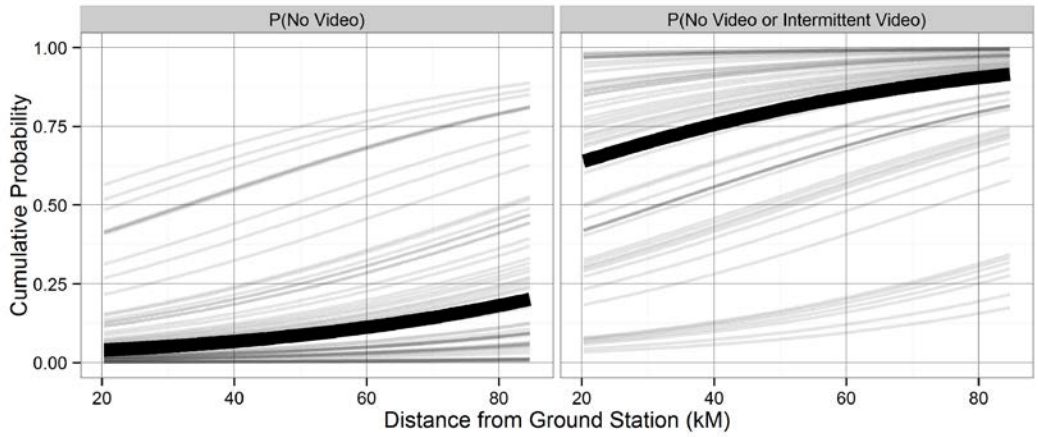

Figure 4: Thresholds between levels of video quality for each mission

6. Conclusions

We have proposed two approaches for analyzing continuously-observed ordinal response variables with continuously-observed piece-wise constant predictors. While intuitively appealing and straightforward to execute, the thinning approach to regularization may not be adequate for autocorrelated data with identifiable functional forms. When autocorrelation is not constant within functions, thinning may overestimate the effective sample size of the data, leading to Type 1 errors when determining factor significance. A more rigorous study of the performance of thinning would be necessary to show this result generally, but this is beyond the scope of our case study. The thinning approach to regularization we proposed leverages the piecewise-constant structure of our data, and we can be used to generate a data set that accurately represents the observed test data. Combined with accurate regularization, cumulative logistic mixed model regression can provide population-level performance estimates and measure mission-to-mission variation.

This approach may be generalizable beyond application for the United State DoD. In clinical trials, patients may be monitored frequently over time, and some medications require administration characterized by heavy doses followed by recovery periods during which the drug is administered in a lower dosage or not at all. A plot over time of the level of medication may resemble a step function. Within the DoD many other unmanned air vehicle systems are in development, many of which will have the ability to transmit full motion video. It is likely that testing these systems will generate data similar to that discussed above. As functional data continues to become increasingly common across all fields, novel analysis approaches must continue to be pursued.

Bibliography

Agresti, A. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc.: New York, NY, 2007; Chapter 6.

Agresti A. *Categorical Data Analysis 3rd Edition*. John Wiley & Sons, Inc.: Hoboken, NJ, 2012; 301-314.

Jupp, D. L. The "Lethargy" theorem—A property of approximation by γ-polynomials. *Journal of Approximation Theory,* 1975; *14*(3), 204-217. DOI: 10.1016/0021-9045(75)90056-8

Picard, F., Robin, S., Lavielle, M., Vaisse, C., & Daudin, J. J. A statistical approach for array CGH data analysis. *BMC* B*ioinformatics*, 2005; 6(1), 27. DOI: 10.1186/1471-2105-6-27

Ramsay, J.O. and Silverman, B.W. *Functional Data Analysis* (2nd Edition), Springer-Verlag: New York, NY, 2005.

Myers, R.J., Montgomery, D.C., Vining, G.G., and Robinson, T.J., *Generalized Linear Models With Applications in Engineering and the Sciences 2nd edition*. John Wiley & Sons, Inc.: New York, NY 2010.

Spiriti, S., Eubank, R., Smith, P. W., & Young, D. Knot selection for least-squares and penalized splines. Journal of Statistical Computation and Simulation, 2013; 83(6), 1020-1036. DOI: 10.1080/00949655.2011.647317

Distance (kM)	Altitude (m)	Video Quality	Mission	Day of Exercise
29.4	1259.42	1	7	4
46.3	2358.69	1	7	4
45.1	2437.39	1	7	4
46.4	2441.11	1	7	4
46	2435.69	1	7	4
53.7	2435.27	1	7	4
55.5	2438.62	1	7	4
⋮	⋮	⋮	⋮	⋮
33	1409.26	2	7	4

Table 1: Data points from example mission selected by thinning.

Distance (kM)	Altitude (m)	Video Quality	Time	Mission	Day of Exercise
45.1	2437	1	47	7	4
53.7	2438	1	58	7	4
50.2	2437	2	13	7	4
53.6	2743	2	163	7	4
41.1	2742	2	39	7	4
30.7	2742	2	126	7	4
32	1864	3	7	7	4
31.3	1417	2	6	7	4

Table 2: Factor values over intervals between change points.

Factor	Parameter Estimate	Standard Error	p-value
Altitude (transformed)	-0.144	0.0858	0.057
Distance (transformed)	-0.292	0.1513	0.029
Date of Test	0.218	0.0653	<0.000

Table 3: Cumulative logit mixed model regression results based on thinning regularization.

Factor	Parameter Estimate	Standard Error	p-value
Altitude (transformed)	-0.109	0.2236	0.628
Distance (transformed)	-0.513	0.2480	0.039
Date of Test	0.207	0.0654	0.002

Table 4: Cumulative logit mixed model regression results using smoothing regularization.